


FEATURE BASED INDEXING OF HAND-WRITTEN TEXT IMAGES

NAGARAJAH KUNTHARSHAN

 This dissertation was submitted to the Department of Computer Science and Engineering of the University of Moratuwa in partial fulfillment of the requirement for the Degree of M Sc in Computer Science

University of Moratuwa, Sri Lanka
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

March 2010

Abstract

Identity management system which is maintaining a manual database system to store its data is facing a big challenge of searching a particular entry from the database if a query rose where the date of birth of the person is not known. Person Identification department, as an example, the existing database is in the form of Pre-designed hand written physical cards with the details. The cards are served separated by sex and ordered by date of birth.

In case of a person whose date of birth is not known, there is a challenge of going through the cards one by one to spot the exact record. So, person registration department decided to computerize the database and found out the following. Since the hand writing is very poor there is no chance of using character recognition software. The better option could be manually enter the data into a database, but quotations submitted for the tender called for this task was much higher than the potential level of the department.

Finally, feature extraction method is found out as an ideal solution for this task. In this approach all the cards are scanned and saved in the system using batch scanning. Each file is pre-processed and the number of characters in the name is saved to the database as index for the corresponding scanned image. The search operation on the database based on the number of characters of the name will list down the name of the possible card and the corresponding card also will be fetched from the saved location and previewed. Among the cards that are identified, user needs to find the cards manually.

This will narrow down the search. System will fail in counting the number of characters in the name if there is no space left between two characters or if one character is split "into two by mistake. To handle these challenges, a new intelligent algorithm should be generated with the ability to understand the order or the pattern of occurrences of the characters and make the decision based on them